Toward Healthy Aging: Temporal Regression for Disability Prediction and Warning Decision-Making

Jianfei Zhang[1(✉)], Lifei Chen[2], and Shengrui Wang[1,2]

[1] Université de Sherbrooke, Sherbrooke, Québec, Canada
jianfei.zhang@usherbrooke.ca
[2] Fujian Normal University, Fuzhou, Fujian, China

Abstract. Virtually all countries in the world are experiencing growth in the number and proportion of seniors in their population. Almost half of these seniors live with one or more disabling conditions. This highlights the concern about when, and how probably, a disability is likely to occur in aging people. In this paper, we mathematicize this concern as a prediction and a warning of the onset of disability. As such, we propose to start by transforming the fitting problem into a series of independent survival learning and prediction problems. Our approach can use all repeated measures of disability-specific factors and, more importantly, effectively quantify their different impacts on the onset of disability. We also present a new approach for estimating the time of onset and determining the better-timed warning of disability onset. To evaluate our time predictions and warning decisions, we develop four evaluation metrics based on the criteria we explore for the aging study. The results of comparative experiments and ablation studies on the elderly cohorts across Canada demonstrate the effectiveness of our approach.

1 Introduction

Each of us is aging. In every country in the world, the population is aging faster than ever before. As of November 2022, the global population aged 65 or over numbers 800 million. Over the next three decades, this number is projected to more than double, reaching over 1.5 billion in 2050 [24]. Statistics Canada said that over 861,000 Canadian people (which is 2.3% of the population) aged 85 and older were counted in the 2021 Census, more than twice the number observed in the 2001 Census [19]. Over the next 25 years (by 2046), the population aged 85 and older could triple to almost 2.5 million people in Canada [19]. Age has a significant impact on health – *the older you get, the more likely you are to become disabled*. The number of people aging with disabilities is on the rise all around the world. According to the 2017 Canadian Survey on Disability [18], roughly 37.8% of Canadians aged 65 and above had been identified as having some form of disability regarding flexibility, dexterity, mobility, cognition, vision, hearing, pain-related, etc. The disability was severe enough to limit them to some extent in their daily activities, especially for those aged 85 and older.

C. Strauss et al. (Eds.): DEXA 2023, LNCS 14147, pp. 378–392, 2023.
https://doi.org/10.1007/978-3-031-39821-6_31

The concern of this work is a generation of better-timed warnings of the onset of disability. The warning in advance induces healthcare providers to take early actions toward healthy aging. For example, they can make early informed decisions regarding healthcare and social services, which will enable the elderly to delay the onset and reduce the severity of disabilities, thereby maintaining optimal health and quality of life. What is needed for this concern is, from a computational perspective, to predict '*when and how probably a disability is likely to occur in aging people*', and determine '*when a warning of disability onset should be issued*'. Hence, we have two tasks: *time-to-onset prediction* – i.e., predicting the time elapsing from the beginning of the follow-up until the onset of disability, and *better-timed warning decision* – i.e., determining the time when the lowest onset-free probability below which the warning should be issued. There is as yet no documented work to make such a time-to-onset prediction and the warning decision based on that prediction. For example, one may be able to estimate that a 65-year-old woman has a 70% probability of living without a disability for one year and only a 10% probability for two years but is incapable of determining the exact time when this woman will be disabled and nailing down the time of warning. Making such a prediction and decision is arduous. Late warnings might lead to conditions' worsening, while too early warnings may increase unnecessary nursing and sometimes make the actions unfeasible. This is the motivation for the work reported here, whose aim is to accurately predict the *onset-free probability over time*, estimate the *time-to-onset*, and determine the *better-timed warning*.

To make better predictions and decisions, we shall examine the link between the onset of disability and various social, demographic, geographic, and economic factors that impact healthy aging, such as general health and well-being, physical activity, and the use of healthcare services. For example, among older adults, injuries due to falls threaten independent living, mobility, and functional ability (such as the ability to engage in regular exercise), and increase the risk for future disabilities [16]. Note, however, that these factors may be repeatedly measured and therefore the measures may vary over time (e.g., taking measurements with an interval of a particular time frame). For example, the seniors may be asked, '*Have you had a fall in the past 6, 12, and 18 months?*' Here, the factor 'fall' is measured repeatedly at 6, 12, and 18 months, producing three measures. Handling these repeated measures has been challenging [6] because of the difficulty in using these measures together to train the existing time-to-event data analysis methods [27]. Besides this, we suffer from *censoring*: most of the older people in our study had not experienced a disability by the close of the study period (e.g., the past 6, 12, or 18 months), and therefore the time of the onset is unknown for a subset of the seniors. *How to build the link between these unknown onset times and repeated measures is the key issue of this work.*

In this paper, we first formalize aging data in two parts: factors' repeated measures and encoded onset times, and then develop temporal survival regression (TR) to explore the link between these two time-dependent parts. It is trained by learning the measures of various factors and the onset of disability over time.

With the learned model, we predict the onset-free probabilities at different times, estimate the time of onset in the future, and determine the better-timed warning of disability. We conduct experiments on real-life aging data and compare the performance of our approach with state-of-the-art models. An extensive ablation study is also performed to investigate the effectiveness of each elaboration of our approach. The experimental results demonstrate that our approach performs better than other baselines and TR's variants, yielding more accurate predictions of onset-free probability and estimates of onset time. We summarize the contributions as follows:

- Our approach is computationally simple: it requires only a sequence of logistic regression fits and the operations are easily understood in terms of regression modeling. We transform the fitting problem into a series of independent learning problems by predicting the encoded responses that we redefined based on the observations. The prediction tasks at different times are independent of each other, but the predicted results are highly related.
- Our approach is conceptually interpretable: it performs predictions and makes prediction-based healthcare decisions on a reasoning basis so that people readily understand how factors are jointly related to form the final predictions and decisions. We impose the impact of the factors and their repeated measures on the onset of disability, where the impact values explored by the model identify the risk or protective effect of both time-varying and time-invariant factors, thereby guiding healthcare actions.
- Our approach is empirically effective: it performs a retrospective study on a Canadian cohort including a bunch of non-disabled people, where the results demonstrate not just the high accuracy achieved by the model but its predictive confidence in dealing with censoring contexts. We develop our own assessment criteria based on the consensuses in life science and mathematicize the criteria as evaluation metrics in the absence of ground truth.

2 Related Work

We will review the work related to time-to-event data analysis. Broadly speaking, time-to-event data analysis can be classified into two main categories: statistical methods and machine-learning-based methods, which share the common goal of predicting the time of the event. Statistical approaches can be grouped into nonparametric (e.g., Nelson-Aalen estimator [1], Kaplan-Meier estimator [9]), parametric (e.g., accelerated failure time model [26]), and semi-parametric estimators (e.g., semi-parametric Cox proportional-hazards model [4]), developed primarily for retrospective cohort studies, each has its inherent disadvantages. Nonparametric approaches are intended to generate unbiased descriptive statistics, but generally cannot be used to assess the effect of multiple factors on the response variable (e.g., time-to-event probability), and the parametric approaches suffer from a critical weakness, relying as they do on the assumption that the underlying failure distribution (i.e., how the probability of failure changes over time) has been correctly specified. For semi-parametric approaches, the assumption about

how the factors influence the risk of failure is often violated in practical use. The increasing availability of a wide variety of data (e.g., time-varying factors) poses more challenges to the statistical approaches and is stimulating numerous research efforts that use machine learning methods in conjunction with time-to-event modeling. For example, feed-forward neural networks have been used for time-to-event data analysis [28]. Although the feed-forward network can preserve most of the advantages of a typical Cox proportional-hazards hypothesis, it was still not the optimal way to model the baseline variations. This was the rationale for deep learning studies [11]. Additionally, recurrent neural networks [29] proposed to compute the survival function by considering a series of binary classification problems, each leading to the estimation of the survival probability in a given interval of time. Besides neural networks, typical examples include multi-task learning [13], Gaussian process [7], active learning [25], transfer learning [12], etc.

In practice, the factors' effects (e.g., the effect of a treatment) may change over time with longer follow-up [8]. To accommodate such situations, there has been a surge of interest in learning time-varying coefficients instead of time-invariant ones. The varying coefficient models are a very important tool to explore the dynamic pattern. The association between repeated measures and the outcome has been modeled in various ways. The work related to repeated measures focuses mainly on the analysis of time-varying risk factors. For the Cox model, [23] estimated time-varying coefficients by maximizing a kernel-weighted partial likelihood, while [22] employed a local empirical partial likelihood smoothing. Time-varying coefficients were also used in [14] to describe the potential time-varying effects of factors on breast cancer. The proportionality assumption may not hold in practice when factor effects change over time.

3 Problem Statement

In our study, each senior is identified by two response variables – i.e., 'censor' $C \in \{0, 1\}$ and 'stamp' time $S \in \mathbb{R}^+$. The response 'censor' $C = 0$ indicates that the time of disability onset, say T, is uncensored, where the disability occurred right at the 'stamp' time S (i.e., the last recorded time), that is, $T = S$. If $C = 1$, the onset time T is censored, and the stamp time S underestimates the true but unknown T, i.e., $T > S$. As an example for three old Canadians aged over 65 shown in Table 1 (e.g., $T = 7$ months for the 77-year-old Québecois), and otherwise $C = 1$ means T is censored (i.e., unknown) because of dropout (e.g., the 69-year-old Ontarian dropped out of the study at 11 months) or early end of study (e.g., the 86-year-old Albertan has been free of disability throughout the 18-month study period). Each senior is described by D factors, such as 'province' and 'age' that are time-invariant and 'depression', 'sleep', and 'fall' that are time-varying (i.e., repeatedly measured). All these factors' measures can be expressed as $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_V) \in \mathbb{R}^{D \times V}$ at V different times $\tau_1 < \cdots < \tau_V$, where $\mathbf{x}_v = (x_{v1}, \ldots, x_{vD})^\top$. For time-invariant factor d, we have $x_{1d} = x_{2d} = \cdots = x_{Vd}$.

Table 1. Factors and responses for three aging Canadians.

Risk Factors							Response	
time-invariant			time-varying				Censor	Stamp
Province	Age	··	Depression	··	Fall	Smoking	(C)	(S)
Québec	77	··	mild	··	no	sometimes	0	7
			mild	··	yes	sometimes		
			..	..	..	..		
			no	··	no	seldom		
Ontario	69	··	severe	··	no	often	1	11
			moderate	··	no	seldom		
			..	..	..	..		
			moderate	··	no	sometimes		
Alberta	86	··	no	··	yes	never	1	18
			no	··	no	never		
			..	..	..	..		
			mild	··	no	never		

To say that someone remains onset-free means that they are still at risk of the onset of disability, that is, the disability has not yet occurred in them. Hence, the onset-free probability at time t means the probability of remaining onset-free up to time t, i.e., the probability that the disability will not occur earlier than time t, that is, $\Pr(T \geq t)$. Our learning task is to find a mapping function $f_{\mathbf{W}}$ ($\mathbf{W}$ is the learnable parameters) for predicting the probabilities at $t_1 < t_2 \cdots < t_K$. Then, based on these predictions, we shall determine the onset time T at which the newly designed error E is lowest and the warning policy p minimizing the newly defined warning cost Q, as follows:

$$\left.\begin{array}{l} \mathbf{x}_1, \ldots, \mathbf{x}_{v:\tau_v \leq t_1} \xrightarrow{f_{\mathbf{W}}} \Pr(T \geq t_1) \\ \qquad\qquad\qquad\qquad \vdots \\ \mathbf{x}_1, \ldots, \mathbf{x}_{v:\tau_v \leq t_K} \xrightarrow{f_{\mathbf{W}}} \Pr(T \geq t_K) \end{array}\right\} \begin{array}{l} \xrightarrow{\arg\min E} T \\ \\ \xrightarrow{\arg\min Q} p \end{array}$$

4 Our Approach

4.1 Encoding

The first thing we need to do is encoding the dualistic response, C and S, as the encoded responses Y, where its value at time t can be given by

$$y[t] = (-\mathbb{1}(C = 1))^{\mathbb{1}(S<t)},$$

which takes the value 1 if the disability happened before or at time t, i.e., $S \geq t$, and otherwise either 0 if $C = 0$ or –1 if $C = 1$. Here, $\mathbb{1}(judgment)$ is an indicator function taking a value of 1 if the *judgment* is true and 0 otherwise. Table 2 shows the encoded response for the three Canadians. Each encoded response $y[t_k]$ indicates whether the disability has occurred by time t_k, being 1 if it has occurred, 0 if it has not occurred, and –1 if unknown. Once $y[t_k]$ becomes "0" it will not turn over to "1". There are thus $K + 1$ legally possible sequences of the form $(1, 1, \ldots, 0, 0, \ldots)$, including the sequences composed of all "1"s and all "0"s. For the 67-year-old Québecois, the encoded response remains "1" until $S = 7$ and "0" thereafter; for the other two, whose onset time is censored, the encoded response is "1" until the stamp time and "–1" thereafter.

Table 2. An example of the encoded responses for three Canadian seniors aged over 60.

Encoded Temporal Responses											Response	
$y[1]$	$\cdot\cdot$	$y[7]$	$y[8]$	$\cdot\cdot$	$y[11]$	$y[12]$	$\cdot\cdot$	$y[18]$	$y[19]$	$\cdot\cdot$	C	S
1	1	1	0	0	0	0	0	0	0	0	0	7
1	1	1	1	1	1	–1	–1	–1	–1	–1	1	11
1	1	1	1	1	1	1	1	1	–1	–1	1	18

4.2 Training

For senior $i \in \mathcal{G}_0 = \{i | \forall i : C_i = 0\}$ with known encoded responses $Y = (y[t_1], \ldots, y[t_K])$ at times $t_1 < \cdot\cdot < t_K$ and measures $\mathbf{X}$ $(\tau_V \leq t_K)$, we can estimate the probability of Y via the generalized logistic regression

$$\Pr(Y \mid \mathbf{X}; \mathbf{W})_0 = \frac{\exp(\mathbf{W} * \mathbf{X} \cdot \boldsymbol{\Delta} \cdot \mathbb{1}(\mathbf{y} \leq 0))}{\mathbf{exp}(\mathbf{W} * \mathbf{X} \cdot \boldsymbol{\Delta} * \mathbf{A}) \cdot \mathbf{1}} \tag{1}$$

$$\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_V) \in \mathbb{R}^{D \times V}$$

$$\mathbf{w}_v = (w_{v1}, \ldots, w_{vD})^\top \in \mathbb{R}^{D \times 1}, \forall v = 1, 2, \ldots, V$$

$$\mathbf{W} * \mathbf{X} = (\mathbf{w}_1 \cdot \mathbf{x}_1, \ldots, \mathbf{w}_V \cdot \mathbf{x}_V) \in \mathbb{R}^{1 \times V}.$$

The regression coefficient $\mathbf{W}$ quantifies how the factors and their repeated measures affect the chance of the senior remaining free of disability, where the coefficients $\mathbf{w}_v$ shows the contribution of D measures at time τ_v. The sum of transformed measures across V time points is given by the column-wise Hadamard product [15]. To quantify the different impacts of the D factors and their V repeated measures on the appearance of Y, we develop the time-decay ratio, which is determined by the elapsed time, as follows,

$$\boldsymbol{\Delta} = \mathbf{exp}(\delta(k, v)) \in \mathbb{R}^{K \times V} \tag{2}$$

$$\delta(k, v) = -(t_k - \tau_v) \times \mathbb{1}(t_k \geq \tau_v). \tag{3}$$

Here, **exp** means the element-wise exponential matrix. The decay ratio $\exp(\sigma(k, v))$ is the exponential of the time difference between t_k and τ_v, indicating that the impact of measure at time τ_v on the probability at time t_k decreases as time elapses. The more time elapses, the more the impact is reduced. For instance, a fall-caused injury poses an effect on the onset and this effect will go down as time goes on. We use the lower triangular identity matrix $\mathbf{A} = (\alpha_1, \ldots, \alpha_K) \in \mathbb{R}^{K \times K}$, where $\alpha_{ij} = 1$ if $i \geq j$ and 0 otherwise, to explore the onset-free probabilities. The denominator means the total score (where $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^{1 \times K}$) of the probability during the period $[t_k, t_{k+1})$.

For senior $i \in \mathcal{G}_1 = \{i | \forall i : C_i = 1\}$ with unknown onset time, the encoded responses before the stamp time are consistent. Hence, we have

$$\Pr\left(Y \mid \mathbf{X}; \mathbf{W}\right)_1 = \frac{\mathbf{exp}(\mathbf{W} * \mathbf{X} \cdot \boldsymbol{\Delta} * \mathbf{A} \cdot \mathbb{1}(\mathbf{y} \leq 0)) \cdot \mathbf{1}}{\mathbf{exp}\left(\mathbf{W} * \mathbf{X} \cdot \boldsymbol{\Delta} * \mathbf{A}\right) \cdot \mathbf{1}}. \tag{4}$$

The numerator is the sum of the risks of the target responses appearing. We learn $\mathbf{W}$ by minimizing the negative logarithm of the above likelihood across all seniors via expectation-maximization [5] with suitable initialization, as follows,

$$\min_{\mathbf{W}} \; P(\mathbf{W}) - \sum\nolimits_{i \in \mathcal{G}_0} \log(\Pr(Y_i | \mathbf{X}_i; \mathbf{W})_0) - \sum\nolimits_{i \in \mathcal{G}_1} \log(\Pr(Y_i | \mathbf{X}_i; \mathbf{W})_1), \tag{5}$$

where the elastic-net penalty $P(\mathbf{W}) = \lambda_1 \|\mathbf{W}\|_1 + \lambda_2 \|\mathbf{W}\|_2^2$ makes the loss function strongly convex and leads to a unique minimum. We choose the hyperparameters λ_1 and λ_2 on an independent validation set.

4.3 Prediction

Given the known measures $\mathbf{X}'$ for a new senior at times $\tau_1 < \cdots < \tau_{V'}$ before time t_k and the learned parameters $\widehat{\mathbf{W}} \in \mathbb{R}^{D \times V}$, predicting the onset-free probability at time t_k is actually predicting the encoded responses $Y'[t_k] = (1, 1, \ldots, 1) \in \mathbb{R}^{1 \times k}$. This probability, $\Pr\left(Y'[t_k] \mid \mathbf{X}'; \widehat{\mathbf{W}}\right)$, can be estimated by the temporal regression shown in Eq. 1. (In what follows, the notation with a '' indicates it is redefined on the test set.)

4.4 Decision-Making

When our model yields the predicted onset-free probabilities at different times for the new senior, we will make two decisions: the time of onset for this senior and the time of warning.

- To estimate the time of onset, we define the relative error at time t as follows:

$$E(t) = \sum\nolimits_{k=1}^{K} (|\log t - \log t_k|) \Pr\left(Y'[t_k] \mid \mathbf{X}'; \widehat{\mathbf{W}}\right),$$

thereby seeking the time at which E is the lowest, i.e., the time of onset is

$$\widehat{T} = \mathrm{argmin}_{t \in \{t_1, \ldots, t_K\}} E(t).$$

- To determine the better-timed warning, we define p_L as the threshold for warning generation, so that a warning is put in place as soon as the predicted probability goes down below this threshold, where L is the ideal lead time between the time of warning and the time of onset T. The better-timed warning can be issued at the time when either this senior becomes disabled or his/her predicted probability becomes lower than the threshold, whichever occurs first, i.e.,

$$\mathcal{T}(p_L) = \min\{T, \inf\{t | \Pr(Y[t]|\mathbf{X}; \widehat{\mathbf{W}}) < p_L\}\}, t \in \{t_1, \ldots, t_K\}.$$

For any decision-maker-determined L, the optimal policy $\widehat{p}_L$ should minimize the total cost of the warning generation, i.e.,

$$\widehat{p}_L = \text{argmin} \sum\nolimits_{i \in \mathcal{G}_0} Q_\alpha(L, T_i - \mathcal{T}_i(p_L)),$$

where $Q_\alpha(L, T_i - \mathcal{T}_i(p_L))$ is the cost of a warning at $T_i - \mathcal{T}_i(p_L)$ days before the actual onset time T_i for senior i. Particularly, $Q_\alpha(L, L)$ (which equals 0) is the cost of the warning at L days early and $Q_\alpha(L, 0)$ at the onset time (i.e., no early warning). We use the pinball loss [21] as the cost in this paper.

5 Experiments

5.1 Data

The Canadian Community Health Survey provided a survey between 2015 and 2017 that focused on the health of Canadians living in the ten provinces by examining various factors that affect healthy aging. In this study, we extracted data from the survey respondents from 3,604 Canadians aged 65 and over. Table 3 (top) presents the statistics of the data. For the factors, we tested for pairwise correlations between independent factors using a correlation matrix; if the Pearson correlation coefficient [10] was greater than 0.75 for two factors, we removed the one we deemed to be of lesser importance. In doing so, we obtained 24 factors, of which 12 are time-invariant and the other 12 are time-varying, see Table 3 (bottom). The numeralization of categorical factors (except the two numerical factors: age and BMI) was implemented through Softmax normalization.

5.2 Baselines

Since there are no methods particularly able to handle repeated measures, we selected as our baselines seven survival models that can deal with survival data a bit like our aging data.

- The nonparametric Kaplan-Meier estimator [9] considers the number of disabled people, m_k, at t_k and non-disabled people, n_k, by time t_k, producing the onset-free probability $\Pr(T \geq t) = \prod_{k:t_k \leq t}(1 - m_k/n_k)$.
- The parametric approach assumes $T \sim \text{Weibull}(\xi, q) = q\xi^q t^{q-1}$, where the scale of the distribution is determined by ξ and the shape by q, and predicts the probability at time t as $\Pr(T \geq t) = \exp(-(\xi t)^q)$.

Table 3. Data statistics (top) and 24 factors (bottom).

Population by	number of people
gender: M / F	2,045 (56.7%) / 1,559 (43.3%)
age group: 65–74 / 75–84 / 85+	2,089 (58.0%) / 1,188 (33.0%) / 327 (9.0%)
health group: disabled / censored	711 (19.7%) / 2,893 (80.3%)
12 time-invariant factors	**12 time-varying factors**
age, gender, BMI, marital, province, education, cognition, dietary supplement, sleep, pain, chronic conditions, alcohol use	caregiving, medication use, oral health, depression, transportation, fall, smoking, social participation, physical activity, care receiving, lifestyle changes, pension

- The semi-parametric Cox regression [4] estimates the probability $\Pr(T \geq t \mid \mathbf{X}) = \exp(-H_0(t)\exp(\beta\mathbf{x}))$. The baseline hazard H_0 is the cumulative hazard with no consideration of the factors and determined by the Breslow's estimator [3] and the coefficient β describes the relationship between outcome and factors. We adopted this model to include a single measure $\mathbf{x}$ at time t.
- MTLR [13] builds multiple independent logistic regressors for prediction. Given the measures $\mathbf{x}_k$ at t_k, it predicts the probability at t_k, i.e., $\Pr(T \geq t_k|\mathbf{X}) = \exp\sum_{k=1}^{K}\mathbf{w}_k\mathbf{x}_k / \sum_{k=0}^{K}\exp(\sum_{l=k+1}^{K}(\mathbf{w}_l\mathbf{x}_l))$.
- DeepHit [11] uses a neural network to predict the probability, by adding all outcomes up, i.e., $\Pr(T \geq t_k|\mathbf{X};\omega) = \sum_{k=1}^{K}\Pr(Y[t_k]|\mathbf{x}_k;\omega)$, where ω is the link between the factors and the onset. We used the single-hit setting here.
- SNN [29] uses neural networks to estimate binary classification scores of remaining onset-free at fixed time intervals, where the network's outcomes are considered as the onset-free probability at different time intervals and scaled by a sigmoid function σ: $\Pr(\tau_k|\mathbf{X};\mathbf{W}) = \sigma(\mathbf{w}_k^{\text{out}} \cdot \sigma(\mathbf{W}^{\text{hide}}\mathbf{X}^{\text{hide}}))$, where the networks' parameters $\mathbf{w}_k^{\text{out}}$ and $\mathbf{W}^{\text{hide}}$ are the weights for the output layer and the hidden layers, respectively.

5.3 Variants

To analyze the significance of our model's settings, we conducted an extensive ablation study, in which we developed our model variants by respectively reducing the time-dependent impact (i.e., $\mathbf{\Delta}$ in Eq. 1), the repeated measures (i.e., $\mathbf{X}$ in Eq. 1), the estimate for censoring seniors (i.e., $\Pr(Y \mid \mathbf{X};\mathbf{W})$ in Eq. 4), and the elastic-net penalty (i.e., $P(\mathbf{W})$ in Eq. 5). The four variants are thus:

- TR-censor, which ignores the seniors with no confirmed onset and does not include Eq. 4 in the objective function when learning model parameters.
- TR-impact, which ignores the time-dependent impact of repeated measures on the outcomes and causes each element of $\mathbf{\Delta}$ to be $\exp 0 = 1$.
- TR-static, which discards repeated measures and uses $\mathbf{W} = (0, \ldots, 0, \mathbf{w}_V) \in \mathbb{R}^{D\times V}$ and $\mathbf{X} = (0, \ldots, 0, \mathbf{x}_V) \in \mathbb{R}^{D\times V}$ to predict the onset-free probability.

- TR-penalty, which excludes the elastic-net penalty $P(\mathbf{W})$ from Eq. 5 and therefore learns the regression coefficients $\mathbf{W}$ without regularization.

5.4 Evaluation

To evaluate the models' predictive power, we consider three questions that a good model should be able to answer well:

- **Is the senior would be likely to be disabled in 18 months?**
- **Which one of the two seniors is more likely to be disabled?**
- **How accurate is the prediction that a senior will be disabled?**

These questions can be answered based on the following facts:

- Seniors who have an onset confirmed during the study period should have a lower onset-free probability at the end than those without an onset.
- Every senior who has an onset should have a smaller onset-free probability at his/her onset time than all those who remain onset-free.
- Seniors who have an onset of a disability at a certain time should have an onset-free probability as close as to 0%, while, for those without the onset, the probability should be approaching 100%.

We formalize these facts to develop three evaluation metrics that can be used to quantify a model's ability to address the questions. The restricted concordance index (rC-index) measures the models' predictive ability at $\mathcal{T}^*$ (i.e., at the end of the study). The unrestricted concordance index (uC-index) is a generalization of rC-index [20] across all comparable pairs $\mathcal{P} = \{(i,j)|\forall i,j : T_i \leq S_j\}$. The censoring mean squared error (C-MSE) is an overall error of predicted probability across all stamp times. They are defined as follows:

$$r\text{C-index} = \frac{1}{|\mathcal{G}_0'| \times |\mathcal{G}_1'|} \sum_{i \in \mathcal{G}_0'} \sum_{j \in \mathcal{G}_1'} \mathbb{1}\left(\Pr(Y_i'[\mathcal{T}^*]|\mathbf{X}_i) < \Pr(Y_j'[\mathcal{T}^*]|\mathbf{X}_j)\right)$$

$$u\text{C-index} = \frac{1}{|\mathcal{P}|} \sum_{i,j \in \mathcal{P}} \mathbb{1}\left(\Pr(Y_i'[T_i]|\mathbf{X}_i) \leq \Pr(Y_j'[S_j]|\mathbf{X}_j)\right)$$

$$\text{C-MSE} = \frac{1}{|\mathcal{G}_0' \cup \mathcal{G}_1'|} \sum\nolimits_{i \in \mathcal{G}_0' \cup \mathcal{G}_1'} \left(C_i - \Pr(Y_i'[S_i]|\mathbf{X}_i)\right)^2.$$

The three metrics are highly independent of each other. This means that a model which performs very well on one may not do well on the other two. A sophisticated prediction model should achieve a high rC-index and uC-index with a low C-MSE. Additionally, to evaluate the predicted onset time, we define the relative error (RE) that measures the difference between the estimated time $\widehat{T}$ and ground truth T for all seniors with a disability in the test set:

$$\text{RE} = \frac{1}{|\mathcal{G}_0'|} \sum\nolimits_{\forall i \in \mathcal{G}_0'} \min\left\{\left|(\widehat{T}_i - T_i)/T_i\right|, 1\right\}.$$

Table 4. The 10 cross-validation rC-index, uC-index, C-MSE, and RE test results, expressed as mean (standard deviation). The best results are in **bold**. There is no definite answer to the ideal lead time; 60-days is deemed reasonable, according to the health care system.

		rC-index	uC-index	C-MSE	RE	avg. Q_α ($L = 60$ days)	
						$\alpha = 0.2$	$\alpha = 0.5$
Baselines	KM	.673(.031)	.669(.028)	.372(.029)	.583(.059)	20.6(2.2)	37.9(3.4)
	Weibull	.687(.027)	.682(.034)	.274(.036)	.410(.094)	32.9(6.5)	24.9(3.3)
	Cox	.693(.024)	.673(.033)	.313(.035)	.542(.073)	28.3(5.4)	26.9(2.5)
	MTLR	.744(.035)	.739(.030)	.197(.021)	.311(.053)	17.9(1.5)	31.7(3.8)
	DeepHit	.735(.032)	.701(.029)	.234(.025)	.392(.085)	23.6(4.3)	28.9(2.2)
	TR	**.757**(.022)	**.753**(.029)	**.192**(.022)	**.308**(.068)	**12.7**(2.5)	**18.3**(3.4)
Variants	TR-censor	.718(.024)	.702(.033)	.269(.022)	.426(.072)	33.9(6.5)	24.5(3.0)
	TR-impact	.730(.037)	.717(.026)	.294(.034)	.465(.068)	17.2(6.6)	29.4(1.7)
	TR-static	.698(.020)	.685(.023)	.293(.033)	.504(.072)	25.7(2.0)	25.8(4.5)
	TR-penalty	.735(.028)	.723(.032)	.238(.019)	.533(.065)	19.5(3.6)	26.4(7.1)

5.5 Results

Table 4 shows the 10 cross-validation results on the test data. Our approach outperforms all the other baselines, achieving an rC-index improvement of 1.3% and an uC-index improvement of over 2% in comparison with the second-best model MTLR. MTLR performs better than other baselines mainly due to its particular sequencing setting [13], which is a bit similar to our approach here. Overall, the machine-learning-based models (MTLR and DeepHit) perform better than the statistical models (KM, Weibull, and Cox), revealing the strong processing capabilities and applicability of multi-task learning and neural networks for dealing with complex aging data. Weibull performs badly partly because the log-logistic distribution assumption does not fit our aging distribution well. In most cases, the accuracy in terms of the uC-index is slightly lower than the rC-index because sorting the onset time of all people accurately in the log-rank test [2] is more challenging. More importantly, our approach achieves not only accurate predictions of the onset-free probability but accurate estimates of onset time, compared to all the baselines. On the other hand, TR consistently outperforms all variants. Specifically, the superior performance of TR relative to TR-censor reveals that censors are also informative for model learning. In addition, the improvement of TR over the TR-impact demonstrates the benefit of the dynamic impact of repeated measures on the outcomes. Moreover, the comparison between TR and the TR-static reveals the importance of considering repeated measures. The superior performance of TR relative to TR-penalty demonstrates the effectiveness of the regularization. By using our TR, the warning can be issued at the lowest cost in the context of a 60-day ideal lead time.

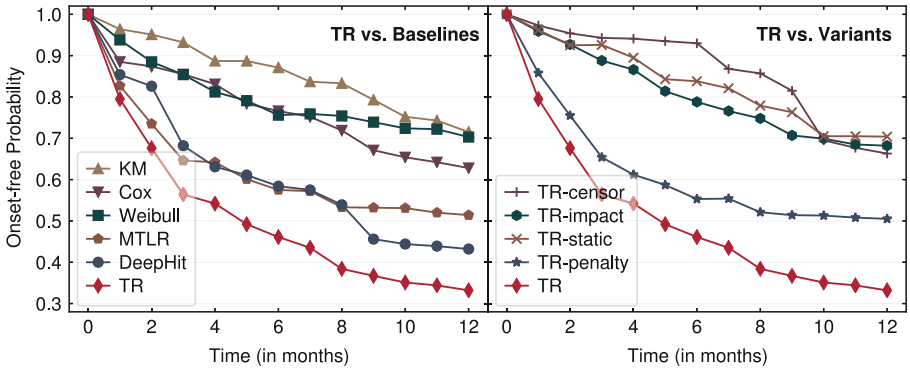

Fig. 1. Average predicted onset-free probability for all disabled seniors.

Predicted Probability. Figure 1 shows the average of the predicted onset-free probability for all disabled seniors through the 12 months. It can be seen that TR yields probabilities that differ highly from other models, where the probability produced by our model at every time interval (e.g. between 4 and 5 months) is much lower than others. The curves produced by the statistical models are close together and much higher than for the other two machine-learning-based models. Among the four variants, TR-penalty produces much lower probabilities than the other three variants.

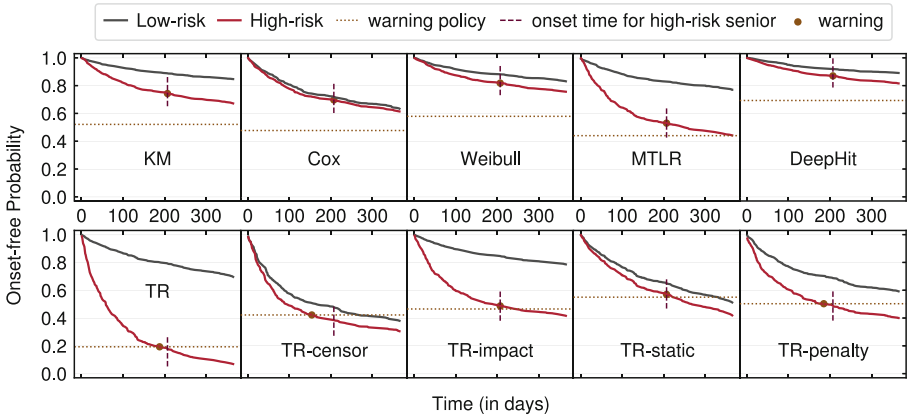

Fig. 2. Our model's predicted onset-free probability for the low-risk and high-risk seniors, estimated onset time for the high-risk senior, and warning policy.

Case Study. For the sake of investigation, Fig. 2 shows the predicted onset-free probabilities for two seniors: a high-risk versus a low-risk senior. (*N.B.*: We consider seniors who experienced a disability during the study period to be at high risk and the others at low risk.) The high-risk senior here is a 69-year-old

woman who became disabled at 207 days (shown by the vertical dashed lines). The low-risk senior is a 75-year-old man who remained onset-free at the end of the study period. It can be seen that only MTLR and TR can clearly distinguish between the two seniors, with TR yielding the largest difference between them. More importantly, TR can predict the extremely low probability for the high-risk senior at 7 months (the onset-free probability is predicted to be lower than 20%). Among these models, only TR, TR-censor, and TR-penalty can generate a warning (shown by the orange point) before the time of onset (i.e., 207 days), where the warning policy (i.e., probability threshold, shown by the orange horizontal dash) yielded by TR is 0.193, 0.422 by TR-censor, and 0.503 by TR-penalty. Note that, the warning issued by our model is 74 days before the disability onset, which is close to the ideal lead time (i.e., 60 days). This high-risk senior could thus be issued a warning and offered advice on early intervention. This is crucial for aging people who are likely to have a severe disability at a specific time.

5.6 Factors' Impact

Figure 3 shows the factors' impact produced by our approach. Here positive impacts reveal a risky effect while negative for a protective effect [17]. The risk factors (e.g., fall and sleep) are associated with a higher likelihood of disability onset, while the protective factors (e.g., lifestyle changes and caregiving) have a cumulative effect on the development of disability.

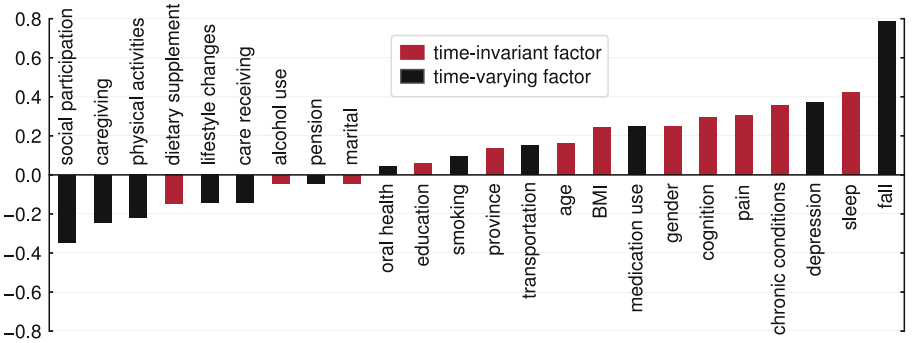

Fig. 3. Factors' impact yielded by our model.

6 Conclusions

This paper proposes a complete paradigm for modeling aging data, predicting onset-free probability, determining the time of onset and better-timed warning, and evaluating the predictions and decisions. The proposed approach successfully addresses the particularities of aging data (i.e., censored responses and repeated

measures) when performing onset prediction and decision-making for a Canadian cohort. It achieves a high prediction accuracy and a low decision error and warning cost, in comparison with various baseline models and our approach's variants. It is the first attempt to develop a machine-learning-based method for predicting older people's possible disabilities, and, of course, we will further improve our approach to address other issues in aging research, e.g., the heterogeneous cohorts (e.g., interprovincial difference) and the time inconsistency (e.g., the difference between measurement times and the times of onset).

Acknowledgments. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Discovery Grant RGPIN-2020-07110 and Discovery Accelerator Supplements Grant RGPAS-2020-00089, the National Natural Science Foundation of China (NSFC) under Grant No. U1805263.

References

1. Aalen, O.: Nonparametric estimation of partial transition probabilities in multiple decrement models. Ann. Stat. 534–545 (1978)
2. Bland, J.M., Altman, D.G.: The logrank test. BMJ **328**(7447), 1073 (2004)
3. Breslow, N.: Covariance analysis of censored survival data. Biometrics 89–99 (1974)
4. Cox, D.R.: Regression models and life tables. J. R. Stat. Soc. Ser. B Stat. Methodol. **34**, 187–220 (1972)
5. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B Stat. Methodol. 1–38 (1977)
6. Everitt, B.S.: The analysis of repeated measures: a practical review with examples. J. R. Stat. Soc. Ser. D Stat. **44**(1), 113–135 (1995)
7. Fernández, T., Rivera, N., Teh, Y.W.: Gaussian processes for survival analysis. In: NeurIPS, pp. 5021–5029 (2016)
8. Fisher, L.D., Lin, D.Y.: Time-dependent covariates in the Cox proportional-hazards regression model. Annu. Rev. Public Health **20**(1), 145–157 (1999)
9. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. J. Amer. Stat. Assoc. **53**(282), 457–481 (1958)
10. Kirch, W. (ed.): Pearson's Correlation Coefficient, pp. 1090–1091. Springer, Netherlands, Dordrecht (2008). https://doi.org/10.1007/978-1-4020-5614-7_2569
11. Lee, C., Zame, W.R., Yoon, J., van der Schaar, M.: DeepHit: a deep learning approach to survival analysis with competing risks. In: AAAI, pp. 2314–2321 (2018)
12. Li, Y., Wang, L., Wang, J., Ye, J., Reddy, C.K.: Transfer learning for survival analysis via efficient l2,1-norm regularized Cox regression. In: ICDM, pp. 231–240 (2017)
13. Lin, H.C., Baracos, V., Greiner, R., Chun-nam, J.Y.: Learning patient-specific cancer survival distributions as a sequence of dependent regressors. In: NeurIPS, pp. 1845–1853 (2011)
14. Liu, M., Lu, W., Shore, R.E., Zeleniuch-Jacquotte, A.: Cox regression model with time-varying coefficients in nested case - control studies. Biostatistics **11**(4), 693–706 (2010)
15. Liu, S., Trenkler, G.: Hadamard, Khatri-Rao, Kronecker and other matrix products. Int. J. Inf. Syst. Sci. **4**(1), 160–177 (2008)

16. Public Health Agence of Canada: Seniors' falls in Canada (2014). https://www.phac-aspc.gc.ca/seniors-aines/publications/public/injury-blessure/seniors_falls-chutes_aines/assets/pdf/seniors_falls-chutes_aines-eng.pdf
17. Seeman, T., Chen, X.: Risk and protective factors for physical functioning in older adults with and without chronic conditions: Macarthur studies of successful aging. J. Gerontol. Ser. B **57**(3), S135–S144 (2002)
18. Statistics Canada: Canadian survey on disability reports (2017). https://www150.statcan.gc.ca/n1/pub/89-654-x/89-654-x2018002-eng.htm
19. Statistics Canada: Census of population (2021). https://www12.statcan.gc.ca/census-recensement/2021/as-sa/index-eng.cfm
20. Steck, H., Krishnapuram, B., Dehing-oberije, C., Lambin, P., Raykar, V.C.: On ranking in survival analysis: bounds on the concordance index. In: NeurIPS, pp. 1209–1216 (2008)
21. Steinwart, I., Christmann, A.: Estimating conditional quantiles with the help of the pinball loss. Bernoulli **17**(1), 211–225 (2011)
22. Sun, Y., Sundaram, R., Zhao, Y.: Empirical likelihood inference for the Cox model with time-dependent coefficients via local partial likelihood. Scand. J. Stat. **36**(3), 444–462 (2009)
23. Tian, L., Zucker, D., Wei, L.: On the Cox model with time-varying regression coefficients. J. Am. Stat. Assoc. **100**(469), 172–183 (2005)
24. United Nations Population Division: World population prospects 2022 (2022). https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf
25. Vinzamuri, B., Li, Y., Reddy, C.K.: Active learning based survival regression for censored data. In: CIKM, pp. 241–250 (2014)
26. Wei, L.J.: The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. Stat. Med. **11**(14–15), 1871–1879 (1992)
27. Wu, L., Liu, W., Yi, G.Y., Huang, Y.: Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. J. Probab. Stat. **2012** (2011)
28. Zhang, J., Chen, L., Ye, Y., Guo, G., Vanasse, A., Wang, S.: Survival neural networks for time-to-event prediction in longitudinal study. Knowl. Inf. Syst. **62**, 3727–3751 (2020). https://doi.org/10.1007/s10115-020-01472-1
29. Zhang, J., Wang, S., Chen, L., Guo, G., Chen, R., Vanasse, A.: Time-dependent survival neural network for remaining useful life prediction. In: Yang, Q., Zhou, Z.-H., Gong, Z., Zhang, M.-L., Huang, S.-J. (eds.) PAKDD 2019. LNCS (LNAI), vol. 11439, pp. 441–452. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-16148-4_34